

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: SPEAR, Patricia G.
MONTGOMERY, Rebecca I.
- (ii) TITLE OF INVENTION: HERPES VIRUS ENTRY RECEPTOR PROTEIN
- (iii) NUMBER OF SEQUENCES: 7
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: DRESSLER, GOLDSMITH, SHORE & MILNAMOW
 - (B) STREET: 180 N. STETSON, SUITE 4700
 - (C) CITY: CHICAGO
 - (D) STATE: ILLINOIS
 - (E) COUNTRY: U.S.A
 - (F) ZIP: 60601
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (vii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: NORTHRUP, THOMAS E.
 - (B) REGISTRATION NUMBER: 33,268
 - (C) REFERENCE/DOCKET NUMBER: XX
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (312) 616-5400
 - (B) TELEFAX: (312) 616-5460
 - (C) TELEX: XX

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1719 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (ix) FEATURE:
 - (A) NAME/KEY: mat_peptide
 - (B) LOCATION: 293..1189
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 293..1192
- (ix) FEATURE:
 - (A) NAME/KEY: sig_peptide
 - (B) LOCATION: 293..406
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCTTCATACC GGCCCTTCCC CTCGGCTTGG .CCTGGACAGC TCTGCCTCCC GCAGGGGCCA 60
 CCTGTGTCCC CCAGCGCCGC TCCACCCAGC AGGCCCTGAGC CCCTCTCTGC TGCCAGACAC 120
 CCCCCTGCTGC CCACCTCTCCT GCTGCTCGGG TTCTGAGGCA CAGCTTGTCA CACCGAGGGG 180
 GATTCTCTTT CTCCTTCTCT TCTGCCCCAC AGCCGACGCA ATGGCGCTGA GTTCCTCTGC 240
 TGGAGTTCAT CCGCTAGCT GGGTTCCCGA GCTGCCGTC TGAGCCTGAG GC ATG 295
 Met
 1
 GAG CCT COT GGA GAC TGG GGG CCT COT CCC TGG AGA TCC ACC CCC AGA 343
 Glu Pro Pro Gly Asp Trp Gly Pro Pro Pro Trp Arg Ser Thr Pro Arg
 5 10 15
 ACC GAC GTC TTG AGG CTG GTG CTG TAT CTC ACC TTC CTG GGA GCC CCC 391
 Thr Asp Val Leu Arg Leu Val Leu Tyr Leu Thr Phe Leu Gly Ala Pro
 20 25 30
 TGC TAC GCC CCA GCT CTG CCG TCC TGC AAG GAG GAC GAG TAC CCA GTG 439
 Cys Tyr Ala Pro Ala Leu Pro Ser Cys Lys Glu Asp Glu Tyr Pro Val
 35 40 45
 GGC TCC GAG TGC TGC CCC ACG TGC AGT CCA GGT TAT CGT GTG AAG GAG 487
 Gly Ser Glu Cys Cys Pro Thr Cys Ser Pro Gly Tyr Arg Val Lys Glu
 50 55 60 65
 GCC TGC GGG GAG CTG ACG GGC ACA GTG TGT GAA CCC TGC CCT CCA GGC 535
 Ala Cys Gly Glu Leu Thr Gly Thr Val Cys Glu Pro Cys Pro Pro Gly
 70 75 80
 ACC TAC ATT GCC CAC CTC AAT GGC CTA AGC AAG TGT CTG CAG TGC CAA 583
 Thr Tyr Ile Ala His Leu Asn Gly Leu Ser Lys Cys Leu Gln Cys Gln
 85 90 95
 ATG TGT GAC CCA GCC ATG GGC CTG CGC GCG ACG CGG AAC TGC TCC AGG 631
 Met Cys Asp Pro Ala Met Gly Leu Arg Ala Thr Arg Asn Cys Ser Arg
 100 105 110
 ACA GAG AAC GCC GTG TGT GGC TGC AGC CCA GGC CAC TTC TGC ATC GTC 679
 Thr Glu Asn Ala Val Cys Gly Cys Ser Pro Gly His Phe Cys Ile Val
 115 120 125
 CAG GAC GGG GAC CAC TGC GCC GGT GCC GCC GTT ACG CCA CCT CCA GCC 727
 Gln Asp Gly Asp His Cys Ala Gly Ala Ala Val Thr Pro Pro Pro Ala
 130 135 140 145
 CGG GCC AGA GGG TGC AGA AGG GAG GCA CCG AGA GTC AGG ACA CCC TGT 775
 Arg Ala Arg Gly Cys Arg Arg Glu Ala Pro Arg Val Arg Thr Pro Cys
 150 155 160
 GTC AGA ACT GCC CCC GGG GAC CTT CTC TCC AAT GGG ACC CTG GAG GAA 823
 Val Arg Thr Ala Pro Gly Asp Leu Leu Ser Asn Gly Thr Leu Glu Glu
 165 170 175
 TGT CAG CAC CAG ACC AAG TGC AGC TGG CTG GTG ACG AAG GCC GGA GCT 871
 Cys Gln His Gln Thr Lys Cys Ser Trp Leu Val Thr Lys Ala Gly Ala
 180 185 190
 GGG ACC AGC AGC TCC CAC TGG GTA TGG TGG TTT CTC TCA GGG AGC CTC 919
 Gly Thr Ser Ser Ser His Trp Val Trp Phe Leu Ser Gly Ser Leu
 195 200 205
 GTC ATC GTC ATT GTT TGC TCC ACA GTT GGC CTA ATC ATA TGT GTG AAA 967
 Val Ile Val Ile Val Cys Ser Thr Val Gly Leu Ile Ile Cys Val Lys
 210 215 220 225
 AGA AGA AAG CCA AGG GGT GAT GTA GTC AAG GTG ATC GTC TCC GTC CAG 1015

Arg Arg Lys Pro Arg Gly Asp Val Val Lys Val Ile Val Ser Val Gln
230 235 240

CGG AAA AGA CAG GAG GCA GAA GGT GAG GCC ACA GTC ATT GAG GCC CTG 1063
Arg Lys Arg Gln Glu Ala Glu Gly Ala Thr Val Ile Glu Ala Leu
245 250 255

CAG GCC CCT CCG GAC GTC ACC ACG GTG GCC GTG AGG AGA CAA TAC CCT 1111
Gln Ala Pro Pro Asp Val Thr Thr Val Ala Val Arg Arg Gln Tyr Pro
260 265 270

CAT TCA CGG GGA GGA GCC CAA ACC ACT GAC CCA CAG ACT CTG GAC CCC 1159
His Ser Arg Gly Gly Ala Gln Thr Thr Asp Pro Gln Thr Leu His Pro
275 280 285

GAC GCC AGA GAT ACC TGG AGC GAC GGC TGC TGA AAGAGGCTGT CCACCTGGCG 1212
Asp Ala Arg Asp Thr Trp Ser Asp Gly Cys
290 295 300

AAACCACCGG AGCCCGGAGG CTTGGGGGCT CCGCCCTGGG CTGGCTCCG TCTCCTCCAG 1272

TGGAGGGAGA GGTGGGGCCC CTGCTGGGCT AGAGCTGGGG ACGCCACGTG CCATTCCCAT 1332

GGGCCAGTGA GGGCCTGGGG CCTCTGTTCT GCTGTGGCCT GAGCTCCCCA GATCCTGTAG 1392

GAGGAGGCC AGTTGCCCT CCGTCACAGA CCACACACCC AGCCCTCCTG GGCCAGCCCA 1452

GAGGCGCCTT CAGACCCAG CTGTCTGCGC GTCTGACTCT TGTGGCTCA GCAGGACAGG 1512

CCCCGGGCAC TGCCTCACAG CCAAGGCTGG ACTGGGTTGG CTGCAGTGTG GTGTTTAGTG 1572

GATACCACAT CGGAAGTGAT TTTCTAAATT GGATTGAAT TCCGGTCTGT TCTTCTATTT 1632

GTGATGAAC AGTGATTGTT GGGAGATGCT GTGGGAGGAT GTAAATATCT TGTTCCTCT 1692

CAAAAAAAAA AAAAAAAAA AAAAAAA 1719

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 299 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Glu Pro Pro Gly Asp Trp Gly Pro Pro Pro Trp Arg Ser Thr Pro
1 5 10 15

Arg Thr Asp Val Leu Arg Leu Val Leu Tyr Leu Thr Phe Leu Gly Ala
20 25 30

Pro Cys Tyr Ala Pro Ala Leu Pro Ser Cys Lys Glu Asp Glu Tyr Pro
35 40 45

Val Gly Ser Glu Cys Cys Pro Thr Cys Ser Pro Gly Tyr Arg Val Lys
50 55 60

Glu Ala Cys Gly Glu Leu Thr Gly Thr Val Cys Glu Pro Cys Pro Pro
65 70 75 80

Gly Thr Tyr Ile Ala His Leu Asn Gly Leu Ser Lys Cys Leu Gln Cys
85 90 95

Gln Met Cys Asp Pro Ala Met Gly Leu Arg Ala Thr Arg Asn Cys Ser
100 105 110

Arg Thr Glu Asn Ala Val Cys Gly Cys Ser Pro Gly His Phe Cys Ile
 115 120 125
 Val Gln Asp Gly Asp His Cys Ala Gly Ala Ala Val Thr Pro Pro Pro
 130 135 140
 Ala Arg Ala Arg Gly Cys Arg Arg Glu Ala Pro Arg Val Arg Thr Pro
 145 150 155 160
 Cys Val Arg Thr Ala Pro Gly Asp Leu Leu Ser Asn Gly Thr Leu Glu
 165 170 175
 Glu Cys Gln His Gln Thr Lys Cys Ser Trp Leu Val Thr Lys Ala Gly
 180 185 190
 Ala Gly Thr Ser Ser Ser His Trp Val Trp Trp Phe Leu Ser Gly Ser
 195 200 205
 Leu Val Ile Val Ile Val Cys Ser Thr Val Gly Leu Ile Ile Cys Val
 210 215 220
 Lys Arg Arg Lys Pro Arg Gly Asp Val Val Lys Val Ile Val Ser Val
 225 230 235 240
 Gln Arg Lys Arg Gln Glu Ala Glu Gly Glu Ala Thr Val Ile Glu Ala
 245 250 255
 Leu Gln Ala Pro Pro Asp Val Thr Thr Val Ala Val Arg Arg Gln Tyr
 260 265 270
 Pro His Ser Arg Gly Gly Ala Gln Thr Thr Asp Pro Gln Thr Leu His
 275 280 285
 Pro Asp Ala Arg Asp Thr Trp Ser Asp Gly Cys
 290 295

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AACCCGGCTC GAGCGCCGC T

21

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAATTCACAC ACACCTAAGG TG

22

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACAAGACCGT TGCACCTC

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4619 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 64..1320

(ix) FEATURE:
 (A) NAME/KEY: mat_peptide
 (B) LOCATION: 64..1317

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AAGCTTGCAT GCCTGCAGGT CCACTCTAGC TGGGTTCCCG AGCTGCCGGT CTGAGCCTGA	60
GGC ATG GAG CCT CCT GGA GAC TGG GGG CCT CCT CCC TGG AGA TCC ACC Met Glu Pro Pro Gly Asp Trp Gly Pro Pro Pro Trp Arg Ser Thr	108
1 5 10 15	
CCC AGA ACC GAC GTC TTG AGG CTG GTG CTG TAT CTC ACC TTC CTG GGA Pro Arg Thr Asp Val Leu Arg Leu Val Leu Tyr Leu Thr Phe Leu Gly	156
20 25 30	
GCC CCC TGC TAC GCC CCA GCT CTG CCG TCC TGC AAG GAG GAC GAG TAC Ala Pro Cys Tyr Ala Pro Ala Leu Pro Ser Cys Lys Glu Asp Glu Tyr	204
35 40 45	
CCA GTG GGC TCC GAG TGC TGC CCC ACG TGC AGT CCA GGT TAT CGT GTG Pro Val Gly Ser Glu Cys Cys Pro Thr Cys Ser Pro Gly Tyr Arg Val	252
50 55 60	
AAG GAG GCC TGC GGG GAG CTG ACG GGC ACA GTG TGT GAA CCC TGC CCT Lys Glu Ala Cys Gly Glu Leu Thr Gly Thr Val Cys Glu Pro Cys Pro	300
65 70 75	
CCA GGC ACC TAC ATT GCC CAC CTC AAT GGC CTA AGC AAG TGT CTG CAG Pro Gly Thr Tyr Ile Ala His Leu Asn Gly Leu Ser Lys Cys Leu Gln	348
80 85 90 95	
TGC CAA ATG TGT GAC CCA GCC ATG GGC CTG CAC GCG ACG CGG AAC TGC Cys Gln Met Cys Asp Pro Ala Met Gly Leu Arg Ala Thr Arg Asn Cys	396
100 105 110	
TCC AGG ACA GAG AAC GCC GTG TGT GGC TGC AGC CCA GGC CAC TTC TGC Ser Arg Thr Glu Asn Ala Val Cys Gly Cys Ser Pro Gly His Phe Cys	444
115 120 125	
ATC GTC CAG GAC GGG GAC CAC TGC GCC GGT GCC GCC GTT ACG CCA CCT Ile Val Gln Asp Gly Asp His Cys Ala Gly Ala Ala Val Thr Pro Pro	492

130						135						140						
CCA Pro	GCC Ala 145	CGG Arg	GCC Ala	AGA Arg	GGG Gly	TGC Cys 150	AGA Arg	AGG Arg	GAG Glu	GCA Ala	CCG Pro 155	AGA Arg	GTC Val	AGG Arg	ACA Thr	540		
CCC Pro 160	TGT Cys	GTC Val	AGA Arg	ACT Thr	GCC Ala 165	CCC Pro	GGG Gly	GAC Asp	CTT Leu	CTC Leu 170	TCC Ser	AAT Asn	GGG Gly	ACC Thr	CTG Leu 175	588		
GAG Glu	GAA Glu	TGT Cys	CAG Gln	CAC His 180	CAG Gln	ACC Thr	AAG Lys	TGC Cys	AGA Arg 185	ATT Ile	CAC His	AAG Lys	ACC Thr	GTT Val 190	GCA Ala Ala	636		
CCC Pro	TCG Ser	ACA Thr	TGC Cys 195	AGC Ser	AAG Lys	CCC Pro	ACG Thr	TGC Cys 200	CCA Pro	CCC Pro	CCT Pro	GAA Glu	CTC Leu 205	CTG Leu	GGG Gly	684		
GGA Gly	CCG Pro	TCT Ser 210	GTC Val	TTC Phe	ATC Ile	TTC Phe	CCC Pro 215	CCA Pro	AAA Lys	CCC Pro	AAG Lys	GAC Asp 220	ACC Thr	CTC Leu	ATG Met	732		
ATC Ile	TCA Ser 225	CGC Arg	ACC Thr	CCC Pro	GAG Glu	GTC Val 230	ACA Thr	TGC Cys	GTG Val	GTG Val	GTG Val 235	GAC Asp	GTG Val	AGC Ser	CAG Gln	780		
GAT Asp 240	GAC Pro	CCC Pro	GAG Glu	GTG Val	CAG Gln 245	TTC Phe	ACA Thr	TGG Trp	TAC Tyr	ATA Ile 250	AAC Asn	AAC Asn	GAG Glu	CAG Gln	GTG Val 255	828		
CGC Arg	ACC Thr	GCC Ala	CGG Arg	CCG Pro 260	CCG Pro	CTA Leu	CGG Arg	GAG Glu	CAG Gln 265	CAG Gln	TTC Phe	AAC Asn	AGC Ser	ACG Thr 270	ATC Ile	876		
CGC Arg	GTG Val	GTC Val	AGC Ser 275	ACC Thr	CTC Leu	CCC Pro	ATC Ile	ACG Thr 280	CAC His	CAG Gln	GAC Asp	TGG Trp	CTG Leu 285	AGG Arg	GGC Gly	924		
AAG Lys	GAG Glu	TTC Phe 290	AAG Lys	TGC Cys	AAA Lys	GTC Val	CAC His 295	AAC Asn	AAG Lys	GCA Ala	CTC Leu	CCG Pro 300	GCC Ala	CCC Pro	ATC Ile	972		
GAG Glu	AAA Lys 305	ACC Thr	ATC Ile	TCC Ser	AAA Lys	GCC Ala 310	GGG Arg	GGG Gly	CAG Gln	CCC Pro	CTG Leu 315	GAG Glu	CCG Pro	AAG Lys	GTC Val	1020		
TAC Tyr 320	ACC Thr	ATG Met	GGC Gly	CCT Pro	CCC Pro 325	CGG Arg	GAG Glu	GAG Glu	CTG Leu	AGC Ser 330	AGC Ser	AGG Arg	TCG Ser	GTC Val 335	AGC Ser	1068		
CTG Leu	ACC Thr	TGC Cys	ATG Met	ATC Ile 340	AAC Asn	GGC Gly	TTC Phe	TAC Tyr	CCT Pro 345	TCC Ser	GAC Asp	ATC Ile	TCG Ser	GTG Val 350	GAG Glu	1116		
TGG Trp	GAG Glu	AAG Lys	AAC Asn 355	GGG Gly	AAG Lys	GCA Ala	GAG Glu	GAC Asp 360	AAC Asn	TAC Tyr	AAG Lys	ACC Thr 365	ACG Thr	CCG Pro	GCC Ala	1164		
GTG Val	CTG Leu	GAC Asp 370	AGC Ser	GAC Asp	GGC Gly	TCC Ser	TAC Tyr 375	TTC Phe	CTC Leu	TAC Tyr	AAC Asn	AAG Lys 380	CTC Leu	TCA Ser	GTG Val	1212		
CCC Pro	ACG Thr 385	AGT Ser	GAG Glu	TGG Trp	CAG Gln	CGG Arg 390	GGC Gly	GAC Asp	GTC Val	TTC Phe	ACC Thr 395	TGC Cys	TCC Ser	GTG Val	ATG Met	1260		
CAC His 400	GAG Glu	GCC Ala	TTG Leu	CAC His	AAC Asn 405	CAC His	TAC Tyr	ACG Thr	CAG Gln	AAG Lys 410	TCC Ser	ATC Ile	TCC Ser	CGC Arg	TCT Ser 415	1308		

CCG GGT AAA TGA GCGCTGTGCC GGCAGCTGC CCTCTCCCT CCCCCCAGC	1360
Pro Gly Lys *	
CCGAGCTGT GCACCCCGCA CACAATAAA GCACCCAGCT CTGCCCTGAA CAGCTTCCGG	1420
TCTCCCTATA GTGAGTCGTA TTAATTTGCA TAAGCCAGCT GCATTAATGA ATCGGCCAAC	1480
GCGCGGGAG AGGCGGTTTG CGTATTGGGC GCTCTTCCGC TTCCTCGCTC ACTGACTCGC	1540
TGCGCTCGGT CGTTCCGGTG CCGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT	1600
TATCCACAGA ATCAGGGGAT AACCGAGGAA AGAACATGTG AGCAAAAGGC CAGCAAAAGG	1660
CCAGGAACCG TAAAAAGGCC CGCTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCCTGACG	1720
AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAGAT	1780
ACGAGCGT TTCCCTCGTA AGCTCCCTCG TGCGCTCTCC TGTTCGACC CTGCGCGTTA	1840
CCGATACCT GTCCGCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT AGCTCACGCT	1900
GTAGGTATCT CAGTTCGGTG TAGTCGTTT GCTCCAAGCT GGGCTGTGTG CACGAACCCC	1960
CCGTTACGCC CGACCGCTGC GCCTTATCCG GTAACATCG TCTTGAGTCC AACCCGGTAA	2020
GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG	2080
TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGGACAG	2140
TATTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT	2200
GATCCGCGAA ACAAAACCACC GCTGGTAGCG GTGGTTTTTT TGTTCGAAG CAGCAGATTA	2260
CAGCGAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC	2320
AGTGAACGA AAACCTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA	2380
CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAATCAAT CTAAAGTATA TATGAGTAAA	2440
CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT	2500
TTCGTTCATC CATAATTGCC TGACTCCCGC TCGTGATGAT AACTACGATA CGGAGGGGCT	2560
TACCATCTGG CCCCAGTGCT GCAATGATAC CCGAGAGACC ACGCTCACCG GCTCCAGATT	2620
TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGTCTCT GCAACTTTAT	2680
CCGCTCCAT CCAGTCTATT AATTGTTGCC GGGGAAGTAG AGTAAGTAGT TCGCCAGTTA	2740
ATAGTTTGGC CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGCTGTTTG	2800
GTATGGCTTC ATTCAGTCC GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCATGT	2860
TGTGCAAAAA AGCGTTAGC TCCTTCGGTC CTCGATCGT TGTCAAGAAT AAGTTGGCCG	2920
CAGTGTATC ACTCATGGT ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG	2980
TAAATGCTTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGATATG	3040
GGCGACCGAG TTGCTCTTGC CCGCGCTCAA TACGGGATAA TACCGCGCA CATAGCAGAA	3100
CITTAAGGT GCTCATATT GGAACGCTT CTTCGGGCG AAAACTCTCA AGGATCTTAC	3160
CGCTGTTGAG ATCCAGTTTG ATGTAACCA CTCGTGCACC CAAGTATCT TCAGCATCTT	3220
TTACTTTTAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAAGG	3280
GAATAAGGGC GACACGGAAT TGTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA	3340

GCATTATATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA	3400
AACAAATAGG GGTTCGCGCG ACATTTCCCG GAAAAGTGCC ACCTGACGTC TAAGAAACCA	3460
TTATTATCAT GACATTAAAC TATAAAAAA GGGGTATCAC GAGGCCCTTT CGTCTCGCGC	3520
GTTTCGGTGA TGACGGTGAA AACCTCTGAC ACATGCAAGCT CCCGGAGACG GTCACAGCTT	3580
GTCTGTAAGC GGATGCCGGG AGCAGACAAG CCCGTCAAGG CCGTCACGCG GGTGTTGGCG	3640
GGTGTGCGGG CTGGCTTAAC TATGCGGCAT CAGAGCAGAT TGTAAGTAGA GTGCACCATA	3700
TCGACGCTCT CCCTTATGCG ACTCCTGCAT TAGGAAGCAG CCCAGTAGTA GGTGAGGCC	3760
GTTGAGCACC GCGGCCGCAA GGAATGGTGC AAGGAGATGG CGCCCAACAG TCCCCGGCC	3820
ACGGGGCGCTG CCACCATACC CACGCCGAAA CAAGCGCTCA TGAGCCGAA GTGGCGAGCC	3880
CGATCTTCCC CATCGGTGAT GTCGCGGATA TAGGCGCCAG CAACCGCACC TGTGGCGCCG	3940
GTGATGCCGG CCACGATGCG TCCGCGTAG AGGATCTGGC TAGTTATTAA TAGTAATCAA	4000
TTACGGGGTC ATTAGTTTAT AGCCCATATA TGGAGTTCCG CGTTACATAA CTTACGGTAA	4060
ATGGCCCGCC TGGCTGACCG CCCAACGACC CCGGCCCAT GACGTCAATA ATGACGTATG	4120
TTCCCATAGT AACGCCAATA GGGACTTTCC ATTGACGTCA ATGGGTGGAC TATTTACGGT	4180
AAACTGCCCA CTGGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCC CCTATTGACG	4240
TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCAGTA CATGACCTTA TGGGACTTTC	4300
CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC CATGGTGATG CGGTTTTGGC	4360
AGTACATCAA TGGCGGTGGA TAGCGGTTTG ACTCACGGGG ATTTCCAAGT CTCACCCCCA	4420
TTGACGTCAA TGGGAGTTTG TTTTGGCACC AAAATCAACG GGACTTTCCA AAATGTCGTA	4480
ACAACTCCCG CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG GTCTATATAA	4540
GCAGAGCTCT CTGGCTAACT AGAGAACCCA CTGCTTAACT GGCTTATCGA AATTAATACG	4600
ACTCACTATA GGGAGACCC	4619

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 418 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met	Glu	Pro	Pro	Gly	Asp	Trp	Gly	Pro	Pro	Pro	Trp	Arg	Ser	Thr	Pro
1				5				10					15		
Arg	Thr	Asp	Val	Leu	Arg	Leu	Val	Leu	Tyr	Leu	Thr	Phe	Leu	Gly	Ala
			20					25					30		
Pro	Cys	Tyr	Ala	Pro	Ala	Leu	Pro	Ser	Cys	Lys	Glu	Asp	Glu	Tyr	Pro
			35					40					45		
Val	Gly	Ser	Glu	Cys	Cys	Pro	Thr	Cys	Ser	Pro	Gly	Tyr	Arg	Val	Lys
			50					55					60		
Glu	Ala	Cys	Gly	Glu	Leu	Thr	Gly	Thr	Val	Cys	Glu	Pro	Cys	Pro	Pro

65 70 75 80

Gly Thr Tyr Ile Ala His Leu Asn Gly Leu Ser Lys Cys Leu Gln Cys
85 90 95

Gln Met Cys Asp Pro Ala Met Gly Leu Arg Ala Thr Arg Asn Cys Ser
100 105 110

Arg Thr Glu Asn Ala Val Cys Gly Cys Ser Pro Gly His Phe Cys Ile
115 120 125

Val Gln Asp Gly Asp His Cys Ala Gly Ala Ala Val Thr Pro Pro Pro
130 135 140

Ala Arg Ala Arg Gly Cys Arg Arg Glu Ala Pro Arg Val Arg Thr Pro
145 150 155 160

Cys Val Arg Thr Ala Pro Gly Asp Leu Leu Ser Asn Gly Thr Leu Glu
165 170 175

Glu Cys Gln His Gln Thr Lys Cys Arg Ile His Lys Thr Val Ala Pro
180 185 190

Ser Thr Cys Ser Lys Pro Thr Cys Pro Pro Pro Glu Leu Leu Gly Gly
195 200 205

Pro Ser Val Phe Ile Phe Pro Pro Lys Pro Lys Asp Thr Leu Met Ile
210 215 220

Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val Ser Gln Asp
225 230 235 240

Asp Pro Glu Val Gln Phe Thr Trp Tyr Ile Asn Asn Glu Gln Val Arg
245 250 255

Thr Ala Arg Pro Pro Leu Arg Glu Gln Gln Phe Asn Ser Thr Ile Arg
260 265 270

Val Val Ser Thr Leu Pro Ile Thr His Gln Asp Trp Leu Arg Gly Lys
275 280 285

Glu Phe Lys Cys Lys Val His Asn Lys Ala Leu Pro Ala Pro Ile Glu
290 295 300

Lys Thr Ile Ser Lys Ala Arg Gly Gln Pro Leu Glu Pro Lys Val Tyr
305 310 315 320

Thr Met Gly Pro Pro Arg Glu Glu Leu Ser Ser Arg Ser Val Ser Leu
325 330 335

Thr Cys Met Ile Asn Gly Phe Tyr Pro Ser Asp Ile Ser Val Glu Trp
340 345 350

Glu Lys Asn Gly Lys Ala Glu Asp Asn Tyr Lys Thr Thr Pro Ala Val
355 360 365

Leu Asp Ser Asp Gly Ser Tyr Phe Leu Tyr Asn Lys Leu Ser Val Pro
370 375 380

Thr Ser Glu Trp Gln Arg Gly Asp Val Phe Thr Cys Ser Val Met His
385 390 395 400

Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Ile Ser Arg Ser Pro
405 410 415

Gly Lys